LC/NAIROBI.

THE ECA STATISTICAL DATA BANK

TABLE OF CONTENTS

1. INTRODUCTION
2. THE ECA STATISTICAL DATA BANK
3. CHARACTERISTICS OF THE LEVEL III AND LEVEL II DATA BANK
4. CONTENTS AND STRUCTURE OF THE LEVEL III DATA BANK
5. MAIN USES OF THE DATA BANK
6. THE ECA STATISTICAL DATA BANK AND ITS MEMBER STATES
7. FUTURE WORK
8. CONCLUSION

1. INTRODUCTION

The ECA Statistical data bank also named PADIS-STAT is one of the components of the Pan African Documentation and Information System (PADIS) project created with the financial support, among others, of the United Nations, the United Nations Development Programme (UNDP), the African Development Bank (ADB), and the Canadian Government through the International Development Research Centre (IDRC). The main objectives of PADIS is to provide access to information required by policy makers, technicians, planners and others engaged in the economic and social development of African States.

The PADIS system will have the following components:

A. PADIS-DEV, the economic and social development information component. A bibliographic reference data base system dealing with all aspects of published and unpublished literature on the development of Africa.

B. PADIS-COM, the set of other complementary bibliographic data bases concerning Africa and to be prepared in collaboration with regional and subregional institutions.

C. PADIS-PROM, the project management component data base concerning development project in their negotiation or implementation phases;

D. PADIS-CORE, the common register data base concerning African specialists, research workers and on-going research activities.

E. PADIS-STAT, the Pan African socio-economic numerical data base which deals with statistical data on Africa.

F. PADIS-TEND, the Pan African technical numerical, non-statistical data bank on national resource, industrial facilities, products and commodities concerning the priority areas of interest of the member countries.

G. PADIS-NET, the telecommunications network for system' interconnection and information dissemination throughout the African continent making available to African users scientific and technical information existing in the industrialized world's data banks by a telecommunications interconnection with the existing international network.

H. PADIS-ADMIN, the integrated system for administrative information including software packages for personnel, finance, stock-keeping and general services management systems.

The ECA statistical data bank is established by the United Nations Economic Commission for Africa (UNECA) to meet the needs of African countries for an up-to-date and [illegible]ly statistical

information and to encourage the use of improved statistics for public purposes and more specifically in the development of regional frame work for statistical information. This objective will be implemented through among others (a) the provision of a Pan African information system using the most recent technologies for data transmission; (b) the establishment of links (interconnexion of systems) with international information networks, data bases and documentation units located outside Africa.

To help in making the statistical data bank operational, a mission organized and financed by the Government of France was asked to under take a feasibility study in a number of African countries and then to consider and submit a set of proposals for organizing the management of the system. An ad-hoc inter-divisional working Group composed of Statistics, Socio-Economic Research and Planning, Population and International Trade, and Finance Divisions was also set up to study the practical problems relating to the implementation of the project. Following the conclusions of this mission and those formulated by the ad-hoc inter-divisional working Group the Executive Secretary of ECA launched in December 1981 the initial regional operations of the ECA statistical data bank. The overall responsibility for the design of the specifications of the ECA statistical data bank and its management has been vested with the Statistics Division. PADIS provides the necessary computer services.

2. THE ECA STATISTICAL DATA BANK

The ECA statistical data bank contains statistical information on social, demographic, economic and environment[illegible]istics of the African states. The major classifications of the contents of the ECA Statistical data bank is contained in Annex 1.

The main objectives of the ECA statistical data bank is to provide a response to the requirements expressed by the African countries concerning the up-to-date information needs of the continent for planning and monitoring economic and social development. The ECA statistical data bank therefore aims to:

A. Compile statistical data in a computerized form;

B. Provide ECA statisticians, economists and demographers with direct and speedy access of [illegible]can statistical data and to provide them with data processing and analysis services;

C. Automate the publications of ECA's st[illegible]istical information system;

D. Disseminate comprehensive and up-to-da[illegible] statistical information to users;

E. Provide assistance service[illegible] to member States who plan to set up statistical data [illegible]

F. Give users other types of services which are required from the system.

The ECA statistical data bank consists of three levels of data bases integrated into one system of information. Level I is expected to provide a comprehensive country profile of one to four pages of statistical data on any one African country or region. It is essentially structural indicators of population social and economic infrastructures, some aggregates of national accounts, and some regional and international comparisons provided over 2 to 5 periods.

Level II contains the bulk of statistical data available at very low level of aggregation over long period of time. It will cover all the fundamental data produced monthly, quarterly and yearly enabling the calculation of dynamic indicators which will allow monitoring of current economic facts.

Level III data base contains selected time series statistical data over a 21 year period for elaborate calculations and econometric modelling. The data contained in this data base are for aggregates and selected commodities which are frequently sought by users.

The implementation of level I depends on the progress of levels II and III. The work of level II is substantial and has been delayed due to the lack of resources, namely adequate computer time and facilities as well as staff.

Figure 1 displays the design concept proposed for the Statistical data bank and consists of seven sub-systems. It is a comprehensive model comprising all the three levels of statistical data bases integrated into one system of information allowing input processing, update processing and output processing. It would be inappropriate at this time to definitely describe each and every element of all the sub-systems; nevertheless a general design is presented for future improvements and expansion.

Sub-systems 1 through 3 are input processing systems for levels II and III data bases while sub-system 5 is input and update processing system for building and maintaining level I data base.

Sub-system 5 is used for direct on-line application. It is also used to update level III data base interactively.

Sub-systems 4 to 6 are used for update processing. Sub-system 6 is an off-line storage and retrieval system of level II data base. It is used to store the data off-line on either magnetic tapes or serial disks. It has the capability of restoring all or selected data from the off-line storage media.

Sub-system 7 is an output processing system for the three data bases. It has the following six major functions:

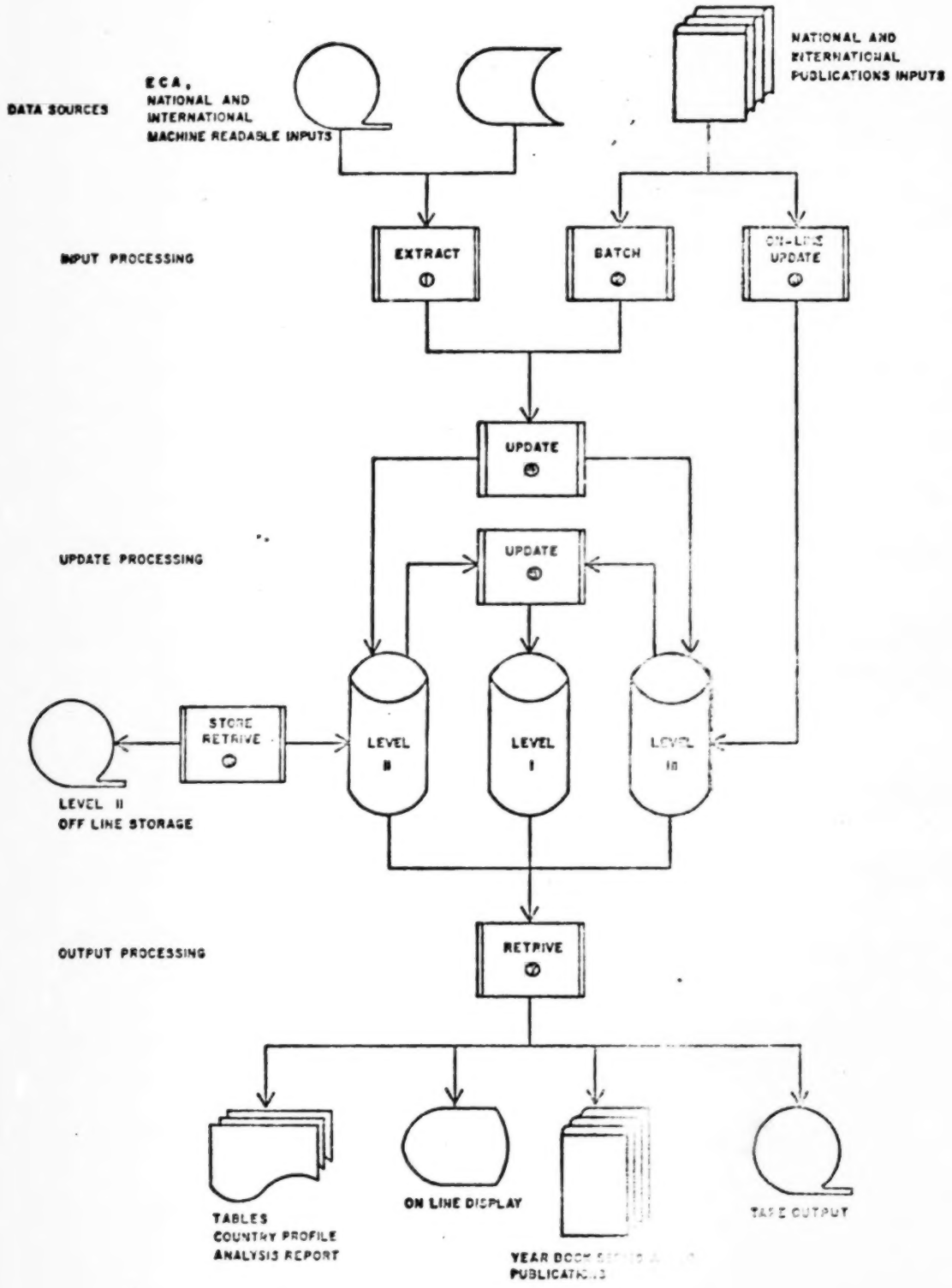

Figure 1. The ECA statistical data bank

A. Produces formatted tables for statistical use.

B. Produces formatted output tables and graphs for the various statistical publications.

C. Produces a comprehensive country profile of one to four pages for a given country or countries on video terminal or gives hard copy printout.

D. Produces an on-line statistical information on video terminals or gives hard copy printout.

E. Produces formatted machine readable data on magnetic tapes or disks.

F. Produces formatted statistical data on magnetic tapes or disks for further analysis.

3. BASIC BASICS OF THE LEVEL III DATA BASE

The level III data base has been developed on HEWLETT PACKARD HP 300 Series III mini computer using the HP Data Base Management System IMAGE 3000 (DBMS). Although business oriented, this DBMS was used because of (a) cost of development of a DBMS in-house would have been prohibitive, and (b) exploration of availability of proprietary packages suitable for statistics and compatible with the existing system has not yet been successful.

User-written programmes have also been developed to bridge the gap between the data base requirements and the facilities available. They can, in particular, interface the system for on-line or batch processing. In addition a user-written table generator package has been developed to produce off-line all the various kinds of tables presently required for ECA Statistics Division's publications from the data base. These programmes are all written in COBOL. For application programmes the data base management system IMAGE 3000 provides a set of library procedures which can be called from BASIC, COBOL, FORTRAN, and SPL.

The data model type of the data base is relational. Every data element can be linked together with every other. That way, information can be retrieved in a variety of different forms with multiple keys. Information can also generally be sorted by any field in the record.

The access to the data base has not, however, been satisfactory owing to an abnormal response time experienced with the computer system. The security system prevents unauthorized persons from deliberately or accidentally gaining access to the data base.

Theoretical aspects and long term options were left for later attention owing to (a) lack of [illegible] sity to computerize as soon as possible ECA Statistics Division's publications using data from the data base.

4. CONTENTS AND STRUCTURE OF THE LEVEL III DATA BASE

The contents of the data base are annual time series statistical data covering the period 1965 to 1985. The data have been entered from pre-processed disk files, data collected for the National and International publications; and from the 1982 African Statistical Yearbook. Provisions have also been made to include in the data base selected monthly, quarterly and half-yearly data for international trade statistics. The data contained in the data base are for aggregates and selected items which are frequently sought by users.

The data relate to the 50 independent countries of Africa which are members of ECA. Furthermore data are also given for Namibia, Reunion , Western Sahara, South Africa, and selected regional groupings.

The capacity of the data base has been set at 150,000 series and can be expanded if disk space resources are made available to the ECA Statistics Division. Over 60,000 time series statistical data have already been entered into the data base. The end-user can also generate his own data from those stored in the data base through an exchange file, irrespective of the language of the application programme available in the set provided by the system.

The series are divided into broad categories of statistics which are namely, population and employment, national accounts, agriculture, industry, transport and communications, finance, price international trade, education and health, and can be expanded to cover such data as environment statistics.

The data base is made up of 9 different sets of data, each containing a specific set of information. The time series statistical data is contained in one detail data set SDATA. Each data record contains 21 values flagged with two characters of footnote and source descriptor, the first of which indicates the type of data (official, estimate, etc.) and the second the source of the data in broad terms.

The other five data sets contain pertinent information on the series identifier, mode of observation, reporter country, partner country and the unit of observation. The elements in these files also serve as keys to dynamically access the detail statistical data set. The data base management system stores pointer information with each detail time series data which links the data record with these five data sets forming a chain.

The remaining three data sets are used for creator identification, password checking, and transaction processing control. The structure and linkage of the data sets are shown in figure 2 below.

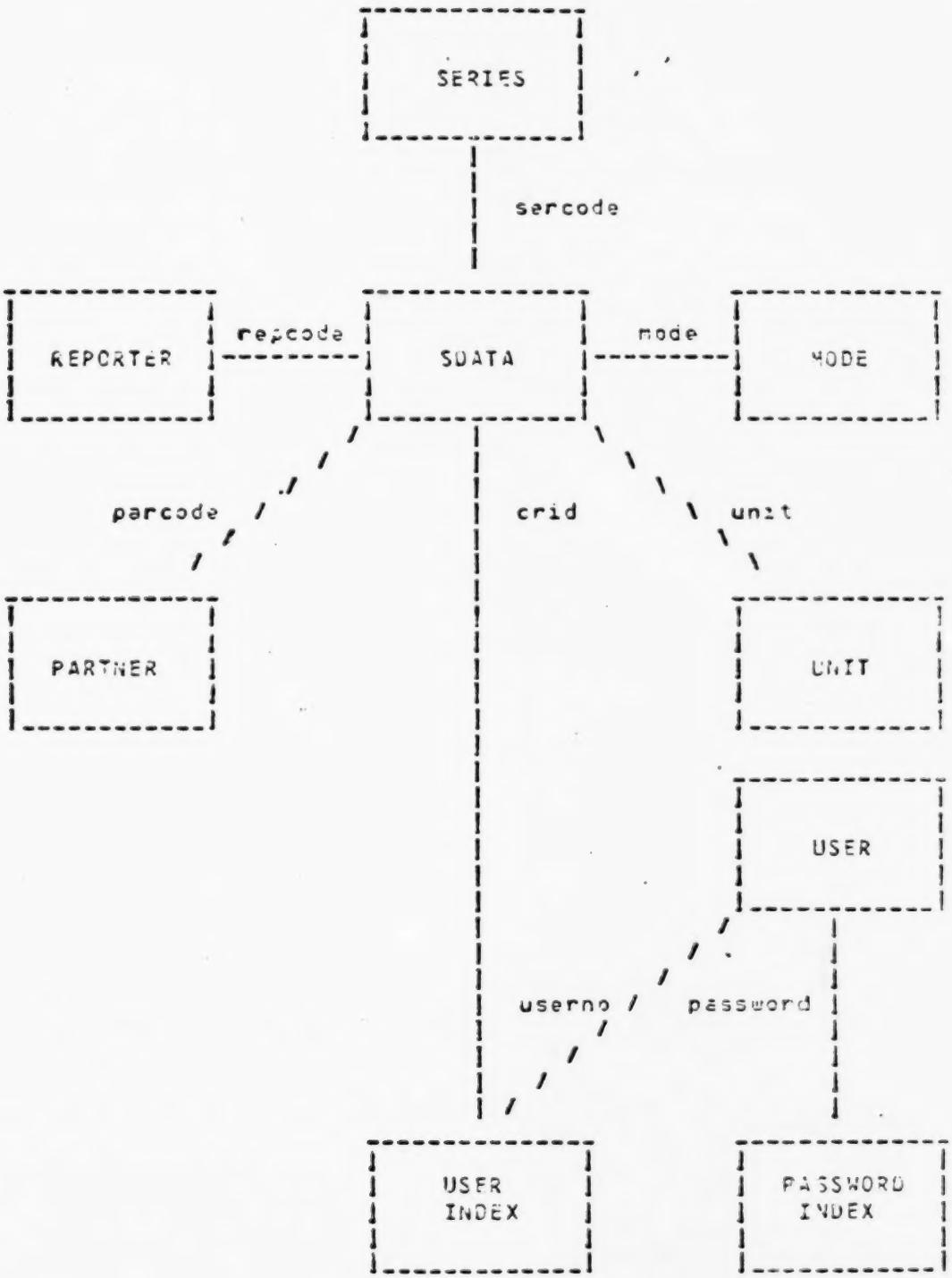

Figure 2. Level III data base structure

5. MAIN USES OF THE DATA BANK

The main functions of the ECA Statistics Division are to provide (a) technical assistance to ECA member countries in statistics, and (b) a regional statistical information service. To perform particularly the second function, the data bank aims primarily at rationalizing the storage of all data which are presently scattered at ECA in various forms.

It will still take some time before the statistical information in the data bank can be made available to enough users through computer readable forms. The preparation of regular and timely statistical publications will therefore be one of the main functions of the data bank for some time. The African Statistical Yearbook parts North Africa and Eastern and Southern Africa have already been so produced and sent for reproduction.

The data base is much more than a single statistical computerized publication. It facilitates calculations, use of econometrics and implementation of models through the use of a more powerful statistical packages such as the Statistical Package for the Social Sciences (SPSS) which can be interfaced with the system. Greater possibility of analysis of the information is, therefore, provided for the solution of and complex problems of social and economic issues of the region.

Bringing the data base to a fully operational stage will enable the ECA Secretariat to perfect and stabilize the methodology that it intends to recommend to its member states. This will also give an opportunity to develop in a joint framework at low cost softwares for calculations, modelling, tabulation, data base management system, and data query packages that the countries could not undertake solely either on account of lack of funds, or because such work is not given a priority. All this joint effort will have a feedback effect on the development of the data banks concerned in the region.

Transborder data flows will be considerably facilitated among member States through the use of the data bank. ECA member countries will thus be able to considerably speed up the dissemination of their social and economic information which was traditionally released slowly, sometimes incompletely, and often with tremendous difficulties through various publications. Each member country will be able to evaluate its own economic and social performance in the light of that of other countries without delay.

The use of the data bank at the regional level will induce training side effects in electronic data processing techniques as well as in statistical documentation that will have to be implemented for each statistical data bank developed. These effects will, no doubt, be of great benefit to statistical development in the region.

The data bank will also facilitate exchange of statistical information between the continent and the international organizations reducing considerably the delay caused by publications production and dispatching. It will, in the long run, make accessible for its African users the data bases available in other parts of the world as supplementary development information sources.

ECA member States will also take advantage of the experience gained by other countries of the world in data base technology. This technology will be for instance first transferred to ECA which will master and if necessary adapt it before releasing it to member states.

Finally, up-to-date statistical materials for socio-economic studies on Africa will be easily available enabling the enrichment of research work at the international level. Complex African socio-economic problems will therefore be more lightened and will be able to receive adequate solutions.

6. IMPACT ON STATISTICAL DATA BANK AND ECA MEMBER STATES

The implementation of the data bank is expected to lead to (a) establishment of statistical data bank in all countries of the region, (b) ready availability of statistical data through telecommunication systems and dissemination on magnetic tapes, hard copies, microfishes, or computer-produced publications at the regional level and in all countries of the region.

To achieve the above goal, the network component PADIS-NET was included in the implementation programme of PADIS. This network component for systems' interconnection and information dissemination throughout the continent has been conceived as a mechanism for the establishment of institutional and technical links between PADIS and the ECA member countries, as well as among the African countries themselves.

The network component of PADIS includes (a) one central co-ordination office sited at ECA Headquarters, (b) four subregional co-ordination centres per ECA sub-regions (North Africa, Central Africa, East and Southern Africa, and West Africa), (c) PADIS institutional participating centres developed in co-operation with regional or subregional specialized organizations in order to co-ordinate their participation in the PADIS programme, and (d) National Information and Documentation Centres (NIDC).

7. FUTURE WORK

The rapid and proper implementation of the data bank requires that adequate resources be made available to the ECA Statistics Division. Despite present insufficient resources, a considerable amount of work has already been accomplished. Certain inputs, however, continue to be lacking for the satisfactory development of the data bank and its release to the member countries. They

are being tackled progressively as the necessary resources become available.

The access to the data base needs to be substantially improved with a view to freeing as much as possible end-users from computer technology knowledge and programming. To this end high-level command language will be developed to facilitate the use of the data base via menus which will prompt at critical places during procedures or allow users to manipulate data with commands that resemble simple English or French.

Presently only simple calculations based on basic arithmetic computations can be used for deriving new statistics. Though the data base does not provide facilities for statistical analysis and graphics, it is possible to generate extract files to be put into statistical analysis and graphic packages. There is therefore a need to further develop the system to offer substantial possibility of calculation, graphs, analysis and modelling which will permit the user to directly input data stored in the data base into his application programme.

The codes of series for commodities have been developed using international classifications. Standard International Trade Classification (SITC), Revision 2, and International Standard Industrial Classification (ISIC) have been used for International Trade and industrial commodities respectively. For the other series the codes have been based on the rough classifications of the UN Directory of International Statistics. The latter does not suit the purpose of coding series for the data base. A particular attention should be given to this matter at the international level in order to standardize codes of statistical series used in international data bases. This would facilitate exchange of information among data banks.

8. CONCLUSION

One of the main objectives which was to give rapid data bank service to users has somewhat been achieved at least for users within the ECA Secretariat. The data base is partly operational in that only data for North, Eastern and Southern African countries have been loaded. Those for the other African sub-regions are underway. A certain amount of work, however, is required before releasing the data bank to outside users.

The project of establishing a regional statistical data bank in Africa is both ambitious and modest. It is ambitious because it aims at providing rapidly accurate statistical information on the social and economic conditions of ECA member States with all tremendous difficulties in the field of statistics. It is modest because only few resources are devoted to it, though it is a very important tool for the socio-economic development of the countries of the region. A non-negligible teething period will therefore be necessary before the data bank becomes fully operational to all potential users.

ANNEX 1

MAJOR CLASSIFICATION OF THE CONTENTS OF THE ECA SUBJECT DATA BANK

1. SOCIAL, DEMOGRAPHIC AND RELATED STATISTICS

1.1 Demographic Statistics
1.2 Housing Statistics
1.3 Labour Statistics
1.4 Education, Science and Technology, Culture and Mass Communication Statistics.
1.5 Health Statistics.
1.6 Distribution of Income, Consumption, and Social Security Statistics
1.7 Other Social, Demographic and Related Statistics.

2. ECONOMIC STATISTICS

2.1 National Accounts and Balances.
2.2 Agriculture, Forestry and Fishery Statistics.
2.3 Industrial and Construction Statistics.
2.4 Trade Statistics.
2.5 Transport, Storage, Communication and Tourism Statistics.
2.6 Finance Statistics.
2.7 Price Statistics.
2.8 Industrial Property Statistics.
2.9 Other Economic Statistic.

3. OTHER STATISTICS

3.1 Environment Statistics.
3.2 Statistics on International Organizations.
3.3 Other Statistics not elsewhere classified.